Verbal Presentation + Transparencies

MARKOV AND THE CREATION OF MARKOV CHAINS

Eugene Seneta

University of Sydney

June, 2006

Preliminary Note.

Pages 1 to 5 were presented *verbally,* accompanied by transparencies of Markov (2) , Nekrasov, Romanovsky, Bernstein, the front cover of the 1906 *Izvestiia (Kazan)* issue, and of Nekrasov's (1902) book.

The electronic picture files together with one of Chuprov, and of the front pages of [36] will also be available on the conference website.

Only some material on Bernstein is not treated more fully in :

Eugene Seneta (2006)*Markov and the the creation of Markov chains.*

In Amy N. Langville and William J. Stewart (Eds.) *MAM2006: Markov Anniversary Meeting.*Boson Books, Raleigh, North Carolina, pp. 1-20.

HISTORICAL INTRODUCTION

Andrei Andreevich Markov was born June 14th (June 2nd, old style), 1856, in Ryazan, Imperial Russia, and died on July 20, 1922, in Petrograd, which was – before the Revolution, and is now again – called Sankt Peterburg (St. Petersburg).

In his academic life, totally associated with St. Petersburg University and the Imperial Academy of Science, he excelled in three mathematical areas: the theory of numbers, mathematical analysis, and probability theory. What are now called Markov chains first appear in his work in a paper of 1906 [27], when Markov was 50 years old. It is the 150th anniversary of his birth, and the 100th anniversary of the appearance of this paper that we celebrate at the Markov Anniversary Meeting, Charleston, South Carolina, June 12 - 14, 2006.

Markov's writings on chains occur within his interest in probability theory. On the departure in 1883 of his mentor, Pafnuty Lvovich Chebyshev (1821 - 1894) from the university, Markov took over the teaching of the course on probability theory and continued to teach it yearly, even in his capacity of a Privat-Dozent (lecturer) after his own retirement from the university as Emeritus Professor in 1905.

The stream of Markov's publications in probability was initially motivated by inadequacies in Chebyshev's treatment of the Central Limit Problem in 1887, and begins with a letter to his friend A.V. Vasiliev (1853 - 1929), which Vasiliev published in the *Izvestiia (Bulletin)* of Kazan University's Physico-Mathematical Society. The Weak Law of Large Numbers (WLLN) and the Central Limit Theorem were the focal probabilistic issues of the times.

The paper [27] in which a Markov chain, as a stochastically dependent sequence for which the WLLN holds, first appeared in Markov's writings, was likewise published in the *Izvestiia.* The paper is motivated by the need to construct a counterexample to what Markov interpreted [50], [57] as a claim in 1902 of P.A. Nekrasov (1853 - 1924) that pairwise independence of summands was necessary as well as sufficient for the WLLN to hold.

Markov was not particularly well-read on the relevant probabilistic literature, and indeed appears not to have been conversant with the fact that the Bernoulli-Laplace urn model, of much earlier provenance, could be cast in the form of a homogeneous Markov chain until, apparently, [35]. When these early models of homogeneous Markov chains are cast in transition matrix form, the transition matrices all have zero entries. Markov, did not completely resolve the matrix structural issues (reducibility, periodicity) which can arise out of such forms of stochastic matrix. His (probabilistic) methodology was strongly focused on the Method of Moments in the guise of conditional and absolute expectations, and double probability generating functions. These functions are closely linked to the determinant [46], [58] and spectral theory of stochastic matrices, and thus necessarily interact with the positioning of zeros in the transition matrix.

His papers on Markov chains utilize the theory of determinants (of finite square matrices), and focus heavily on what are in effect finite stochastic matrices. However, explicit formulation and treatment in terms of matrix multiplication, properties of powers of stochastic matrices, and more generally of inhomogeneous products of stochastic matrices, and of associated spectral theory, are somewhat hidden, even though striking results, rediscovered by other authors many years later, follow from ideas in [27].

The theory of finite non-negative matrices was beginning to emerge only contemporaneously with Markov's [27], [29] first papers on Markov chains, with the work of Perron [44] and Frobenius [14], [15]. The connection between the two directions, Markov and Perron-Frobenius is probably due to von Mises [40]. The theory of finite Markov chains was then developed from this standpoint in the treatises of Fréchet [13] and Romanovsky [45] on homogeneous finite Markov chains.

In our own times, the heavily influential book on finite homogeneous Markov chains has been that of Kemeny and Snell [23], which, while heavily matrix theoretic in its operations, avoids any mention of spectral theory, and in its discussion of ergodicity is closest in spirit to Markov's original memoir [27].

Our mathematical focus is an exploration of the contractivity ideas of that paper in the context of finite stochastic matrices , and specifically of the structure and usage of the Markov-Dobrushin coefficient of ergodicity.

Markov was promoted to Ordinary (full) Professor in 1893, and elected Ordinary (full) Academician in 1896. His mentor Chebyshev had died in 1894. The year 1903 saw the birth of his son, also Andrei, and thus also Andrei Andreevich Markov (1903 - 1979), who himself was to become an eminent mathematician, and Corresponding Member of the Academy of Sciences of the U.S.S.R. The identical 3-part name with his father has sometimes confused western writers producing photographs.

1900 saw the publication of the first edition of Markov's textbook *Ischislenie Veroiatnostei (The Calculus of Probabilities)*. The second edition appeared in Russian in 1908 and was translated into German by Heinrich Liebmann in 1912 as *Wahrscheinlichkeitsrechnung*, which became well-known in the West. The third edition of 1913, substantially expanded, and with a portrait of Jacob Bernoulli, was timed to appear in the year of the 200th anniversary of Jacob Bernoulli's WLLN. Markov organized a Commemorative Meeting of the Academy of Science in honor of the anniversary. Other speakers were Vasiliev and Chuprov. The fourth edition of 1924 [36] was posthumous, and published in Moscow in the early years of the Soviet era. It is again much expanded and, of the Russian-language editions, now the most readily available.

1 Markov's legacy

Tributes to Markov soon after his death appeared in the 1923 *Izvestiia* of the Russian Academy of Science, by Ya.V.Uspensky (= J.V. Uspensky) [65] and A.S. Bezikovich (= A.S. Besicovitch) [5]. Besicovitch [6] also wrote a biographical sketch for the posthumous edition of Markov's monograph [36] .

Yakov Viktorovich Uspensky (1883 - 1947) Uspensky in May 1913 was Privat-Dozent at St. Petersburg University, and translated the celebrated 4th part of Jacob Bernoulli's *Ars Conjectandi* from Latin into Russian, for the 200th anniversary celebrations of Bernoulli's WLLN. Uspensky was Academician of the Russian Academy of Science from 1921. At this time Uspensky was (full) Professor at the University.

Uspensky wrote a tribute to Markov in 1923, soon after Markov's death.

Uspensky's apparently last paper in Russian was published in 1924 in the *Doklady ANSSSR* and is on a probabilistic topic. He appears to have left Leningrad at about this time, and made his way to the United States. His first paper in English, according to Math. Sci. Net., appeared in the *American Mathematical Monthly* in 1927 and was also on a probabilistic topic. A note in [3], p. 167, says he worked in the United States from 1929. In the United States he used the English version

James of Yakov (which is, more accurately translated, as Jacob). Although he continued to write in several areas, and gained considerable distinction, it is largely for his book of 1937, *Introduction to Mathematical Probability*, written as Professor of Mathematics at Stanford University, and based on his lectures there, that he is best known. The book [66] discusses only two-state Markov chains within its chapter *Fundamental Limit Theorems.* It is certainly heavily influenced by the work of the St. Petersburg School of Probability, and specifically by Markov, on the Central Limit Problem. Uspensky's book seems to have brought analytical probability, in the St. Petersburg tradition, to the United States, where it remained a primary probabilistic source until the appearance of W. Feller's *An Introduction to Probability Theory and Its Applications* in 1951. Feller's book contains a great deal on Markov chains, specifically the case of a denumerable number of states, for which a matrix/spectral approach is not adequate, and renewal theoretic arguments are employed.

The matrix method for finite Markov chains was exposited very much from Markov's post-1906 standpoint, in monograph form in Russian, by Romanovsky [45]. It reappeared in English translation by the author of this paper in 1970.

Vsevolod Ivanovich Romanovsky (1879 - 1954)

From its beginning stages in 1918, till his death he was heavily involved in teaching and research in mathematics, and in administration at what became Tashkent State University (initially called Central Asian University). In the early period of his research he worked on differential equations, algebraic questions, and (as expected from his student days at St. Petersburg) on number theory. His later research activities were very largely devoted to the theory and applications of probability theory and mathematical statistics. There was a fundamental paper on finite Markov chains in *Acta Mathematica* in 1936, presumably in imitation of A.A. Markov's French-language publication in this journal in 1910. The distance of Tashkent from St. Petersburg-Petrograd-Leningrad on the other hand, would have worked against any personal contact with A.A. Markov in the last decade or so of Markov's life. Romanovsky's most important scientific work was on finite Markov chains (it began in 1928), and on their generalization. His *magnum opus* of 1949 on this topic [45], however, was algebraically intricate, and received little attention in comparison with the theory of denumerable chains developed by Kolmogorov from the 1930's.

Sergei Natanovich Bernstein (1880- 1968)

In the English-speaking world, finite homogeneous Markov chain theory was reborn with Kemeny and Snell's book.

2 Contractivity principles in Markov's reasoning

In Sections 2.1 to 2.3 we present what may be extracted in essence from Markov (1906), specifically its Section 5.

2.1 Markov's Contraction Inequality

Lemma 1 below, states what we call Markov's Contraction Inequality, which is sometimes inappropriately attributed to Paz [42],

Lemma 1. If $\delta = \{\delta_s\}, \mathbf{w} = \{w_s\}$, are real-valued column N-vectors, and $\delta^T \mathbf{1} = 0$, then

$$
\begin{aligned}
|\delta^T \mathbf{w}| &\leq (\max w_s - \min w_s)\, \frac{1}{2} \sum_{s=1}^{N} |\delta_s| \\
&= \max_{h,h'} |w_h - w_{h'}|\, \frac{1}{2} \sum_{s=1}^{N} |\delta_s|. \qquad \diamondsuit
\end{aligned}
$$

2.2 Contractive property of a stochastic matrix.

The following lemma expresses the averaging property of a stochastic matrix.

Lemma 2. Let $\mathbf{P} = \{p_{ij}\}, i, j = 1, \cdots, N$ be a stochastic matrix, so that $\mathbf{P} \geq \mathbf{0}$, $\mathbf{P1} = \mathbf{1}$. Let $\mathbf{w} = \{w_i\}$ be a real-valued column N-vector, and put

$$\mathbf{v} = \mathbf{P}\ \mathbf{w}$$

Then, writing $\mathbf{v} = \{v_i\}$,

$$\max_{h,h'} |v_h - v_{h'}| \leq H \ \max_{j,j'} \ |w_j - w_{j'}|$$

where

$$H = \frac{1}{2} \max_{i,j} \sum_{s=1}^{N} \ |p_{is} - p_{js}|,$$

so $0 \leq H \leq 1.$ $\diamondsuit$

Lemma 3. Putting $\mathbf{P}^{n-1} = \{p_{sr}^{(n-1)}\}, n \geq 1$, with $\mathbf{P}^0 = \mathbf{I}$ (the unit matrix),

$$\max_{h,h'} |p_{hr}^{(n)} - p_{h'r}^{(n)}| \ \leq H^n, \ n \geq 0. \qquad \diamondsuit$$

If $\mathbf{P} > \mathbf{0}$ i.e. all entries are positive, as Markov effectively assumes, it is clear that $H < 1$ from the expression (for H ; and this is also clearly true if $\mathbf{P}$ has a strictly positive column (i.e. is a "Markov" matrix).

When $H < 1$, the Lemma 3 implies that as $n \to \infty$ all rows of $\mathbf{P}^n$ tend to coincidence (this property was later called "weak ergodicity").

Lemma 4. For fixed r, $\max_h\ p_{hr}^{(n)}$ is non-increasing with increasing n; and $\min_h\ p_{hr}^{(n)}$ is non-decreasing with increasing n, so (since both sequences are bounded) both have limits as $n \to \infty$. When $H < 1$, all rows of $\mathbf{P}^n$ tend to the same limiting probability vector. $\diamondsuit$

The property of all rows of $\mathbf{P}^n$ tending to the same limiting probability distribution came to be called "strong ergodicity".

Notice that the argument of Lemma 4 uses the "backward" form: $\mathbf{P}^n = \mathbf{P}\,\mathbf{P}^{n-1}$; and obtains ergodicity of a finite homogeneous Markov chain, it would seem, at geometric rate of convergence providing $H < 1$, without use of Perron-Frobenius theory of non-negative matrices.

The notation "H" of Lemma 2 is actually Markov's [27], and the expression for it implicitly appears in his paper. The expression for H has been ascribed to Dobrushin [12].

It is appropriate to call H the Markov-Dobrushin coefficient of ergodicity.

2.3 Attribution

A great deal of theory for stochastic matrices/Markov chains can be developed from the inequality of Lemma 1. It remains true if $\mathbf{w}$ is replaced by an N-vector $\mathbf{z} = \{z_i\}$ each of whose elements may be real or complex, so that

$$|\delta^T \mathbf{z}| \leq \max_{h,h'} |z_h - z_{h'}| \, \frac{1}{2} \sum_{s=1}^{N} |\delta_s|$$

This inequality for finite N is due to Alpin and Gabassov [1], where it is proved by induction on N. It follows also from a problem, given without solution in Paz ([42], p.73, Problem 16), and restated in [48], p.583, where the inequality is derived from it. The inequality does not appear in [42].

We propose that the name Markov's Contraction Inequality be used for both the inequality of Lemma 1. and Alpin and Gabassov's result, although the name Lemma PS, as used in the body of Kirkland, Neumann and Shader [24] is a reasonable compromise.

3 Some direct consequences of Markov's contractivity principles

With little extra effort, Lemmas 1 - 4 may be used to obtain direct results, which in qualitative nature are as good as known results using more elaborate (albeit related) superstructure. We give two examples.

3.1 Weak and strong ergodicity of inhomogeneous products

For an $N \times N$ stochastic matrix $\mathbf{P} = \{p_{ij}\}$, write

$$\tau_1(\mathbf{P}) = \frac{1}{2} \max_{i,j} \sum_{k=1}^{N} |p_{ik} - p_{jk}|,$$

3.2 Rate of convergence to ergodicity. The Google matrix

Markov's argument embodied in Lemmas 3 and 4 gives a geometric rate H to equalization of rows,providing $H \equiv \tau_1(\mathbf{P}) < 1$. This is easily extended to the more conventional concept of convergence at geometric rate.

Theorem 1. Suppose $\mathbf{P}$ is $(N \times N)$ stochastic, with $H < 1$, and suppose $\pi^T = \{\pi_r\}$ is the common probability distribution to which each row of $\mathbf{P}^n$ converges as $n \to \infty$. Then for fixed r

$$|p_{ir}^{(n)} - \pi_r| \leq H^n, \ n \geq 0. \quad \diamondsuit$$

The *Google matrix* ([26], p.5) $\mathbf{P}$ is of form

$$\mathbf{P} = \alpha \mathbf{S} + (1 - \alpha)\mathbf{1}\mathbf{v}^T$$

where $0 < \alpha < 1$ and $\mathbf{v}^T > \mathbf{0}^T$ is a probability vector, and both can be arbitrarily chosen. $\mathbf{S}$ is a stochastic matrix. Now

$$H = \tau_1(\mathbf{P}) = \alpha\tau_1(\mathbf{S}) \leq \alpha$$

since $\tau_1(\mathbf{S}) \leq 1$. By Theorem 1, the rate of convergence to the limit distribution vector π^T by the power method is rapid [26], even with a small $(1 - \alpha)$. Using the relations $\pi^T = \pi^T\mathbf{P}, \ \pi^T\mathbf{1} = 1$ it follows immediately that

$$\pi^T = (1 - \alpha)\mathbf{v}^T(\mathbf{I} - \alpha\mathbf{S})^{-1}.$$

4 Measuring Sensitivity under perturbation

4.1 The setting. Norms and bounds

For $\mathbf{x}^T \in \Re^N$, if $\| \cdot \|$ is a vector norm on $\Re^N$, then the corresponding matrix norm for an $(N \times N)$ matrix $\mathbf{B} = \{b_{ij}\}$ is defined by

$$\| \mathbf{B} \| = \sup\{\mathbf{x}^T, \| \mathbf{x}^T \| = 1 : \| \mathbf{x}^T \mathbf{B} \|\}.$$

We focus on the l_p norms on $\Re^N$, where $\| \mathbf{x}^T \|_p = \left(\Sigma_{i=1}^N |x_i|^p\right)^{1/p}$ in the cases $p = 1, \infty$, where $\| \mathbf{x}^T \|_\infty = \max_i |x_i|$. Then

$$\| \mathbf{B} \|_1 = \max_i \sum_{j=1}^N |b_{ij}|, \ \| \mathbf{B} \|_\infty = \max_j \sum_{i=1}^N |b_{ij}|.$$

For any $(N \times N)$ matrix $\mathbf{B} = \{b_{ij}\}$ we may define a Markov-Dobrushin-type coefficient of ergodicity more generally by

$$\tau_1(\mathbf{B}) = \sup\{\delta^T, \| \delta^T \|_1 = 1, \ \delta^T \mathbf{1} = 0 : \ \| \delta^T \mathbf{B} \|_1\}$$

whence [51]

$$\tau_1(\mathbf{B}) = \frac{1}{2} \max_{i,j} \sum_{s=1}^N |b_{is} - b_{js}|.$$

THE FUNDAMENTAL EQUALITY

Suppose $\mathbf{P} = \{p_{ij}\}$ is an $N \times N$ stochastic matrix containing a single irreducible set of indices, so that there is a unique stationary distribution vector $\pi^T = \{\pi_i\}$, $[\pi^T(\mathbf{I} - \mathbf{P}) = \mathbf{0}^T, \pi^T \mathbf{1} = 1]$. Let $\overline{\mathbf{P}}$ be any other $(N \times N)$ stochastic matrix with this structure (the irreducible sets need not coincide), and $\overline{\pi}^T = \{\overline{\pi}_i\}$ its unique stationary distribution vector. Under the assumption on $\mathbf{P}$ the corresponding *fundamental matrix* [23] $\mathbf{Z}$ exists, where $\mathbf{Z} = (I - \mathbf{P} + \mathbf{1}\pi^T)^{-1}$. Set $\mathbf{E} = \{e_{ij}\} = \overline{\mathbf{P}} - \mathbf{P}$. Suppose that there exists an $N \times N$ matrix $\mathbf{C} = \{c_{ij}\}$ such that:

$$\overline{\pi}^T - \pi^T = \overline{\pi}^T \mathbf{E}\,\mathbf{C} \tag{1}$$

Theorem 2 Under our prior conditions on P and $\overline{P}$, and assuming (1) holds:

$$\| \overline{\pi}^T - \pi^T \|_1 \leq \tau_1(\mathbf{C}) \| \mathbf{E} \|_1 \tag{2}$$

$$\| \overline{\pi}^T - \pi^T \|_\infty \leq \mathcal{T}(\mathbf{C}) \| E \|_1 \tag{3}$$

where

$$\mathcal{T}(\mathbf{C}) = \frac{1}{2} \max_j \left(\max_{k,k'} |c_{kj} - c_{k'j}| \right). \tag{4}$$

$\diamondsuit$

The result (1) and (2) was obtained by the author [53] Theorem 2, in the case $\mathbf{C} = \mathbf{C}(\mathbf{u}, \mathbf{v})$ where:

$$\begin{aligned} \mathbf{C}(\mathbf{u}, \mathbf{v}) &= (\mathbf{I} - \mathbf{P} + \mathbf{1}\,\mathbf{u}^T)^{-1} - \mathbf{1}\mathbf{v}^T \\ &= \mathbf{Z} - \frac{\mathbf{1}(\mathbf{u} - \pi)^T \mathbf{Z}}{\mathbf{u}^T\,\mathbf{1}} - \mathbf{1}\,\mathbf{v}^T \end{aligned} \tag{5}$$

for any (real) $\mathbf{v}$, and any (real $\mathbf{u}$) such that $\mathbf{u}^T\,\mathbf{1} \neq 0$. Notice that $\mathbf{C}(\pi, \pi) = \mathbf{A}^\sharp = \mathbf{Z} - \mathbf{1}\,\pi^T$, the *group generalized inverse* [39] $\mathbf{A}^\sharp$ of $\mathbf{A} = \mathbf{I} - \mathbf{P}$; while $\mathbf{C}(\pi, \mathbf{0}) = \mathbf{Z}$.

In fact it is shown in [53] that $\tau_1(\mathbf{C}(\mathbf{u}, \mathbf{v})) = \tau_1(\mathbf{A}^\sharp) = \tau_1(\mathbf{Z}) = \tau_1((\mathbf{I} - \mathbf{P} + \mathbf{1}\mathbf{u}^\mathbf{T})^{-\mathbf{1}})$.

USE OF MARKOV'S CONTRACTION INEQUALITY.

The steps in the proof of (3) are due to Kirkland, Neumann and Shader [24], Theorem 2.2, in the case $\mathbf{C} = \mathbf{A}^{\sharp}$, A key ingredient in the guise of "Lemma PS", is Markov's Contraction Inequality.

Cho and Meyer [11] have shown that the bound on the right of (3) in a different guise also occurs in 1984 in Haviv and Van der Heyden [17], and later, in Cho and Meyer [10]. Haviv and Van der Heyden [17] also used Lemma 1, and Hunter [20], in the proof of his Theorem 3.2, ascribes its result to these authors.

4.2 Measuring sensitivity

The relative effect on π^T of the perturbation $\mathbf{E}$ to $\mathbf{P}$ is measured in a natural way by the quantity

$$\frac{\| \overline{\pi}^T - \pi^T \| / \| \pi^T \|}{\| \mathbf{E} \| / \| \mathbf{P} \|}. \tag{6}$$

From (2), using $\| \cdot \|_1$, and taking $\mathbf{C} = \mathbf{A}^\sharp$ in (1), we see since $\| \mathbf{P} \|_1 = 1$ that (6) $\leq \tau_1(\mathbf{A}^\sharp)$, so $\tau_1(\mathbf{A}^\sharp)$ is a natural condition number to measure the relative sensitivity of π^T to perturbation of $\mathbf{P}$.

Cho and Meyer [11] survey various condition numbers $\kappa_l, l = 1, \cdots, 8$ which have occurred in the literature which satisfy

$$\| \overline{\pi}^T - \pi^T \|_p \leq \kappa_l \| \mathbf{E} \|_q$$

where $(p, q) = (1, 1)$ or $(\infty, 1)$ depending on l . *In this sense,* in particular from their Section 4 $(p, q) = (1, 1)$, $\kappa_6 = \tau_1(\mathbf{A}^\sharp)$, and for $(p, q) = (\infty, 1)$, $\kappa_3 = \kappa_8 = \mathcal{T}(\mathbf{A}^\sharp)$ are condition numbers.

However, one might argue that, inasmuch as a condition number should bound (6), the same norm should be used for numerator and denominator of the left-hand side of (6). This is not the case in expressing (3) in form (6).

In their Remark 4.1, Cho and Meyer [11], p.148, point out, in order to obtain a fair comparison between the bounding tightness of $\kappa_6 = \tau_1(\mathbf{A}^\sharp)$ and $\kappa_3 \equiv \kappa_8 = \mathcal{T}(\mathbf{A}^\sharp)$ that $(\overline{\pi}^T - \pi^T)\mathbf{1} = 0$, so $\| \overline{\pi}^T - \pi^T \|_\infty \leq (1/2)\| \overline{\pi}^T - \pi \|_1$. Hence

$$\kappa_3 = \mathcal{T}(\mathbf{A}^\sharp) \leq \frac{1}{2}\tau_1(\mathbf{A}^\sharp) = \frac{1}{2}\kappa_6,$$

from which they conclude that κ_3 is the tighter condition number. However, one might argue that

$$\sum_j |\overline{\pi}_j - \pi_j| \leq \frac{1}{2}\sum_j \left(\max_{h,h'} |c_{hj} - c_{h'j}| \right) \max_k \| \mathbf{E}_k^T \|_1,$$

and since

$$\tau_1(\mathbf{C}) \leq \Delta(\mathbf{C}) \equiv \frac{1}{2}\sum_j \left(\max_{h,h'} |c_{hj} - c_{h'j}| \right),$$

it follows that the consequent bound on $\| \overline{\pi}^T - \pi \|_1$ is not as tight as when using $\tau_1(\mathbf{A}^\sharp)$.

In their role as condition numbers, $\tau_1(\mathbf{A}^\sharp)$ and $\mathcal{T}(\mathbf{A}^\sharp)$ are not really directly comparable in regard to size, since different versions of the norm $\| \overline{\pi}^T - \pi \|$ are involved.

4.3 Recent related results

The discussion of Section 4 has revolved around (1). Hunter [19] Theorems 2.1 - 2.2 has obtained this equality by using the general form $\mathbf{G}$ of the g-inverse of $\mathbf{I}-\mathbf{P}$:

$$\mathbf{G}=[\mathbf{I}-\mathbf{P}+\mathbf{t}\mathbf{u}^T]^{-1}-\mathbf{1}\mathbf{v}^T+\mathbf{g}\pi^T \tag{7}$$

for any real $\mathbf{t},\mathbf{v},\mathbf{u},\mathbf{g}$ satisfying $\pi^T\mathbf{t}\neq 0$, $\mathbf{u}^T\mathbf{1}\neq 0$, by showing

$$\overline{\pi}^T-\pi^T=\overline{\pi}^T\mathbf{E}\mathbf{G}(\mathbf{I}-\mathbf{1}\pi^T) \tag{8}$$

and that $\mathbf{E}\mathbf{C}(\mathbf{u},\mathbf{v})$ is a special case of $\mathbf{E}\mathbf{G}(\mathbf{I}-\mathbf{1}\pi^T)$.

Equation (8) leads to a bound of general appearance :

$$\| \overline{\pi}^T-\pi^T \|_1 \leq \tau_1(\mathbf{G}(\mathbf{I}-\mathbf{1}\pi^T)) \| \mathbf{E} \|_1. \tag{9}$$

It would seem plausible that $\tau_1(\mathbf{G}(\mathbf{I}-\mathbf{1}\pi^T))$ may give a tighter bound, for some parameter vectors, than $\tau_1(\mathbf{A}^\sharp)=\tau_1(\mathbf{Z})$, but the author has shown that in fact all these values are the same, namely $\tau_1(\mathbf{A}^\sharp)$.

Hunter [20], Corollary 5.1.1, derives the bound

$$\| \overline{\pi}^T - \pi^T \|_1 \leq \mathrm{tr}(\mathbf{A}^\sharp) \| \mathbf{E} \|_1. \tag{10}$$

This bound is not as strict as (2), with $\mathbf{C} = \tau_1(\mathbf{A}^\sharp)$, since [56], p.165, (10), states that

$$\tau_1(\mathbf{A}^\sharp) \leq \mathrm{tr}(\mathbf{A}^\sharp). \tag{11}$$

Work on these issues by J. Hunter and the author is in progress. For the time being, the bound [53]:

$$\| \overline{\pi}^T - \pi^T \|_1 \leq \tau_1(\mathbf{A}^\sharp) \| \mathbf{E} \|_1 \tag{12}$$

remains sharp for the norm used.

References

[1] Yu. A. Alpin and N.Z. Gabassov, *A remark on the problem of localization of the eigenvalues of real matrices (in Russian)*, Izv. Vyssh. Uchebn. Zaved. Matematika, 11(1976), pp. 98–100.

[2] G.P. Basharin, A.N. Langville and V.A. Naumov, *The life and work of A.A. Markov*, Linear Algebra Appl., 386(2004), pp. 3–26.[This was in the Special Issue, pp. 1- 408, on the *Conference on the Numerical Solution of Markov Chains*, 2003.]

[3] J. Bernoulli, *O Zakone Bol'shikh Chisel (On the Law of Large Numbers)*. Nauka, Moscow, 1986.

[4] S. N. Bernstein, *Sobranie Sochinenii: Tom IV [Collected Works, Vol. IV], Teoriia Veroiatnostei i Matematicheskaia Statistika [1911–1946]*, Nauka, Moscow, 1964.

[5] A.S. Bezikovich [Besicovitch], *A. A. Markov's contributions to probability theory (in Russian)*, Izvestiia Rossiiskoi Akademii Nauk, 17(1923), pp. 45–52.

[6] A.S. Bezikovich [Besicovitch], *Biograficheskii Ocherk. [Biographical Sketch]*, [36], pp.III – XIV.

[7] H. Bruns, *Wahrscheinlichkeitsrechnung und Kollektivmasslehre*, Teubner, Leipzig, 1906. [See esp. Achtzehnte Vorlesung].

[8] J.C. Burkill, *Abram Samoilovitch Besicovitch. 1891-1970*, Biographical Memoirs of Fellows of the Royal Society, 17(1971), pp. 1–16.

[9] S. Chatterjee and E. Seneta, *Towards consensus: some convergence theorems on repeated averaging*, J. Appl. Prob., 14(1977), pp. 89–97.

[10] G.E. Cho and C.D. Meyer, *Markov chain sensitivity measured by mean first passage time*, Linear Algebra Appl., 316(2000), pp. 21–28.

[11] G.E. Cho and C.D. Meyer, *Comparison of perturbation bounds for the stationary distribution of a Markov chain*, Linear Algebra Appl., 335(2001), pp. 137–150.

[12] R.L. Dobrushin, *Central limit theorem for non-stationary Markov chains, I and II*, Theory Prob. Appl., 1(1956), pp. 65–80, 329–83 [English translation].

[13] M. Fréchet, *Recherches théoriques modernes sur le calcul des probabilités. Secnd livre. Méthode des fonctions arbitraires. Théorie des événements en chaîne dans le cas d'un nombre fini d'états possibles.* Gauthier-Villars, Paris, 1938.

[14] G. Frobenius, *Über Matrizen aus positiven Elementen*, S.-B. Preuss. Akad. Wiss.,(Berlin), (1908), pp. 471–476; (1909), pp. 514–518.

[15] G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, S.-B. Preuss. Akad. Wiss.,(Berlin), (1912), pp. 456–477.

[16] S. Ia. Grodzensky, *Andrei Andreevich Markov 1856–1922*. Nauka, Moscow, 1987.

[17] M. Haviv and L. Van der Heyden, *Perturbation bounds for the stationary probabilities of a finite Markov chain*, Adv. Appl. Prob., 16(1984), pp. 804–818.

[18] C.C. Heyde and E. Seneta, eds., *Statisticians of the Centuries*, Springer, New York, 2001.

[19] J.J. Hunter, *Stationary distributions and mean first passage times of perturbed Markov chains*, Linear Algebra Appl., 410(2005), pp. 217–243.

[20] J.J. Hunter, *Mixing times with applications to perturbed Markov chains*, LInear Algebra Appl., (2006) to appear.

[21] M. Iosifescu, *Finite Markov Processes and Their Applications*, Wiley, Chichester; Editura Tehnica, Bucharest, 1980.

[22] D.L. Isaacson and R.W. Madsen, *Markov Chains*, Wiley, New York, 1976.

[23] J.G. Kemeny and J.L. Snell, *Finite Markov Chain*, Van Nostrand, Princeton N.J., 1960.

[24] S.J. Kirkland. M. Neumann and B.L. Shader, *Applications of Paz's inequality to perturbation bounds for Markov chains*, Linear Algebra Appl., 268(1998), pp. 183–196.

[25] A.N. Kolmogorov and A. P. Yushkevich, eds., *Mathematics in the 19th Century; Mathematical Logic, Algebra, Number Theory, Probability Theory*, Birkhäuser, Basel, Boston, 1992.

[26] A.N. Langville and C.D. Meyer, *The use of linear algebra by web search engines*, IMAGE, 33(2004), pp. 2–5.

[27] A.A. Markov, *Extension of the law of large numbers to dependent quantities (in Russian)*, Izvestiia Fiz.-Matem. Obsch. Kazan Univ., (2nd Ser.), 15(1906), pp. 135–156 [Also [37], pp. 339–61].

[28] A.A. Markov, *The extension of the law of large numbers onto quantities depending on each other*, [62], pp. 143–158. Translation into English of [27].

[29] A.A. Markov, *Investigation of a notable instance of dependent trials (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 1(3)(1907), pp.61-80. [Also in [37], pp. 363–397.]

[30] A.A. Markov, *Extension of limit theorems of the calculus of probabilities to sums of quantities associated into a chain(in Russian)*, Zapiski Akad. Nauk (St. Petersburg), Fiz.-Mat. Otd., (7th Ser.). 22(9)(1908), 29 pp. [Also [37], pp. 363–397.]

[31] A.A. Markov, *The extension of the limit theorems of the calculus of probability onto a sum of magnitudes connected into a chain*, [62], pp. 159–178. Transalation into English of [30].

[32] A.A. Markov, *Investigation of the general case of trials associated into a chain [in Russian]*, Zapiski. Akad. Nauk (St. Petersburg) Fiz.-Matem. Otd., (8th Ser.) 25(3)(1910), 33 pp. [Also in [37]. pp. 465–507.]

[33] A.A. Markov, *An investigation of the general case of trials connected into a chain*, [62], pp. 181–203. Translation into English of [32] .

[34] A.A. Markov, *On dependent trials not forming a true chain (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 5(2)(1911), pp.113—126. [Also in [37], pp. 399–416.]

[35] A.A. Markov, *On a problem of Laplace (in Russian)*, Izvestiia Akad. Nauk (St. Petersburg), (6th Ser.), 9(2)(1915), pp. 87–104. [Also in [37], pp. 549–571.]

[36] A.A. Markov, *Ischislenie Veroiatnostei (Calculus of Probabilities)*, Fourth (posthumous) edn., Gosudarstvennoie Izdatelstvo, Moscow, 1924.

[37] A.A. Markov, *Izbrannie Trudy (Selected Works)*, A.N.S.S.S.R., Leningrad, 1951.

[38] A.A. Markov [Jr.], *The Biography of A. A. Markov.* In [62], pp. 242-256. Translation into English of [37], pp. 599–613.

[39] C.D. Meyer, *The role of the generalized inverse in the theory of finite Markov chains*, SIAM Rev., 17(1975), pp. 443–464.

[40] R. von Mises, *Wahscheinlichkeitsrechnung*, Fr. Deuticke, Leipzig and Vienna,1931.

[41] Kh.O. Ondar (Ed.), *The Correspondence Between A.A. Markov and A.A. Chuprov on the Theory of Probability and Mathematical Statistics*, [Translated by C. & M. Stein], Springer, New York. 1981.

[42] A. Paz, *Introduction to Probabilistic Automata*, Academic Press, New York,1971.

[43] U. Paŭn, *A class of ergodicity coefficients, and applications*, Math. Reports (Bucharest), 4(2002), pp. 225–232.

[44] O. Perron, *Zur Theorie der Matrices*, Math. Ann., 64(1907), pp. 248–263.

[45] V.I. Romanovsky, *Diskretnie Tsepi Markova*, Gostekhizdat, Moscow, 1949. [English translation by E. Seneta, *Discrete Markov Chains*, Wolters-Noordhoff, Groningen, 1970].

[46] H. Schneider, *The concepts of irreducuibility and full indecomposability of a matrix in the works of Frobenius, König and Markov*, Linear Algebra Appl., 18(1977), pp. 139–162.

[47] P.J. Schweitzer, *Perturbation theory and finite Markov chains*, J. Appl. Prob., 5(1968), pp. 401–413.

[48] E. Seneta, *Coefficients of ergodicity: structure and applications.* Adv. Appl. Prob., 11(1979), pp.576–590.

[49] E. Seneta, *Non-negative Matrices and Markov Chains*, Second edn., Springer, New York, 1981, 2006(revised printing).

[50] E. Seneta, *The central limit problem and linear least squares in pre-revolutionary Russia: The background*, Mathematical Scientist, 9(1984), pp. 37–77.

[51] E. Seneta, *Explicit forms for ergodicity coefficients and spectrum localization*, Linear Algebra. Appl., 60(1984), pp. 187–197.

[52] E. Seneta, *Perturbation of the stationary distribution measured by ergodicity coefficients*, Adv. Appl. Prob., 20(1988), pp. 228–230.

[53] E. Seneta, *Sensitivity analysis, ergodicity coefficients, and rank-one updates for finite Markov chains*. In [63], pp. 121–129.

[54] E. Seneta, *Applications of ergodicity coefficients to homogeneous Markov chains*, Contemporary Mathematics, 149(1993), pp. 189–199.

[55] E. Seneta, *Ergodicity for products of infinite stochastic matrices*, J. Theoret. Probab., 6(1993), pp. 345–362.

[56] E. Seneta, *Sensitivity of finite Markov chains under perturbation,* Statist. Probab. Lett., 17(1993), pp. 163 – 168.

[57] E. Seneta, *Markov and the birth of chain dependence theory*, International Statistical Review, 64(1996), pp. 255–263.

[58] E. Seneta, *Complementation in stochastic matrices and the GTH algorithm*, SIAM J. Matrix Anal. Appl., 19(1998), pp. 556–563.

[59] E. Seneta, *Andrei Andreevich Markov*, In [18], pp.243–247.

[60] E. Seneta, *Statistical regularity and free will: L.A.J. Quetelet and P.A. Nekrasov.* International Statistical Review, 71(2003), pp. 319–334.

[61] O.B.Sheynin, *A.A. Markov's work on probability.* Archive Hist. Exact Sci., 39(1989), pp. 337–77.

[62] O.B.Sheynin, ed., *Probability and Statistics. Russian Papers.*, Selected and Translated by Oscar Sheynin. NG Verlag, Berlin, 2004.

[63] W.J. Stewart, ed., *Numerical Solution of Markov Chains*, Marcel Dekker, New York, 1991.

[64] C.P. Tan. *Coefficients of ergodicity with respect to vector norms.* J. Appl. Prob., 20(1983), pp. 277–287.

[65] J.V. Uspensky, *A sketch of the scientific work of A. A. Markov (in Russian)*, Izvestiia Rossiskoi Akademii Nauk, 17(1923), pp. 19–34.

[66] J.V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.